



Machine Learning Based Interaction Technique Selection For 3D User Interfaces

Jérémy Lacoche, Thierry Duval, Bruno Arnaldi, Eric Maisel, Jérôme Royan

► To cite this version:

Jérémy Lacoche, Thierry Duval, Bruno Arnaldi, Eric Maisel, Jérôme Royan. Machine Learning Based Interaction Technique Selection For 3D User Interfaces. EuroVR 2019 - 16th EuroVR International Conference, Oct 2019, Tallinn, Estonia. pp.33-51, 10.1007/978-3-030-31908-3_3 . hal-02292434

HAL Id: hal-02292434

<https://hal.science/hal-02292434>

Submitted on 19 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Machine Learning Based Interaction Technique Selection For 3D User Interfaces

Jérémy Lacoche^{1,2}, Thierry Duval^{3,4}, Bruno Arnaldi^{5,6}, Eric Maisel^{4,7}, and Jérôme Royan²

¹ Orange Labs, Rennes, France

² IRT b<>com, Rennes, France

³ IMT ATlantique, Brest, France

⁴ Lab-STICC, UMR CNRS 6285, Brest, France

⁵ IriSa, UMR CNRS 6074, Rennes, France

⁶ INSA de Rennes, Rennes, France

⁷ ENIB, Brest, France

Abstract. A 3D user interface can be adapted in multiple ways according to each user's needs, skills and preferences. Such adaptation can consist in changing the user interface layout or its interaction techniques. Personalization systems which are based on user models can automatically determine the configuration of a 3D user interface in order to fit a particular user. In this paper, we propose to explore the use of machine learning in order to propose a 3D selection interaction technique adapted to a target user. To do so, we built a dataset with 51 users on a simple selection application in which we recorded each user profile, his/her results to a 2D Fitts Law based pre-test and his/her preferences and performances on this application for three different interaction techniques. Our machine learning algorithm based on Support Vector Machines (SVMs) trained on this dataset proposes the most adapted interaction technique according to the user profile or his/her result to the 2D selection pre-test. Our results suggest the interest of our approach for personalizing a 3D user interface according to the target user but it would require a larger dataset in order to increase the confidence about the proposed adaptations.

1 Introduction

Today, there is a growing interest in 3D User Interfaces (3DUIs) thanks to the availability of high-quality Virtual Reality (VR) and Augmented Reality (AR) devices for the consumer market. For example, we can cite the Oculus Rift⁸, the HTC Vive⁹ and the Microsoft Hololens¹⁰. A 3D user interface uses a set of interaction techniques in order to achieve a set of high-level tasks. For 3D interactive systems, three kinds of interaction techniques are proposed by Hand [10]: objects manipulation, navigation and application control. In each category, there are a lot of possible techniques that may perform differently depending on the context. In most cases, the designer or the

⁸ <https://www.oculus.com/>

⁹ <https://www.vive.com>

¹⁰ <https://www.microsoft.com/hololens>

developer chooses the interaction technique that he thinks will fit the best to the task. This choice does not depend on the target user.

All users do not have the same interaction habits, the same skills, the same needs. Therefore, some solutions propose to adapt the user interfaces according to the user properties. In that case, we talk about personalization. Using such mechanisms can improve the usability of an application and its attractiveness. Such systems are based on user modeling which aims to build a description of each user (profile, interests, monitoring information) [11]. Personalization can be applied at different levels of granularity, from the choice of the displayed content, its 3D layout [7] or of an interaction technique [15] to the modification of parameters of individual 3D elements [7] [6] and interaction techniques [8]. In this work, we focus on the choice of adapted interaction techniques but our approach could be extended to other kinds of personalization.

In this paper, we explore the use of supervised machine learning in order to infer the most suited interaction technique according to the user profile or to his/her results to a 2D selection pre-test. This profile and the results are part of our user model. Supervised machine learning aims to infer a function from labeled training data in order to predict unseen points [17]. Today, machine learning algorithms are already deployed by recommender systems [23] for the creation of personalized user interfaces in the field of e-commerce, social networks, and search engines. We want to validate the interest of such an approach for 3DUIs for the choice of a selection interaction technique that will be adapted to a new user. For that purpose, we propose to build a dataset with multiple users on a simple selection application in which we will record each user profile, his/her results to a 2D Fitts Law based pre-test and his/her preferences and performances on this application for three different interaction techniques. From this dataset we propose to train Support Vector Machines (SVMs) [1] in order to find the classification function that associates a user model to his/her most adapted selection interaction technique.

Our goal is to be able to integrate our prediction models in frameworks for the creation of adaptive 3D user interfaces. For instance, with the integration of a user model, it could be used seamlessly in the CATHI framework [15] or with the D3PART model [14] in order to compute the suitability score of interaction components.

This paper is structured as follows: first in Section 2 we provide a literature review of personalization systems for interactive systems. Then, in Section 3 we explain the process of creation of the dataset which is the base of our machine learning approach. A prior analysis of the results obtained during this dataset creation is given in Section 4. To continue, our machine learning approach and its results are detailed in Section 5. To finish, these results are discussed in Section 6.

2 Related Work

2.1 User Modeling

The first step to being able to adapt an application to a user is to extract or learn from him/her a profile that contains his/her main characteristics that differentiate him/her from the other users. This is the goal of user modeling. A lot of research work has been published about how to construct user models. Surveys of user modeling are given by

Kobsa [11] [13] and by Fischer [9]. According to Kobsa [12], a user model is defined as a set of information and assumptions on an individual user or on user groups that are used in an adaptation process. From these reviews, we can extract different kinds of properties that can be included in a user model:

- **The user profile.** As explained by Calvary [3], it can include general information such as his/her age his/her gender, his/her size. It can also include information about his/her skills, his/her profession or his/her knowledge.
- **The user preferences and interests.** It can include information about the interaction techniques and the devices that the user prefers and also his/her global or application-oriented interests.
- **The user environment.** It can refer to the physical and social properties of his/her environment [3]. For instance, it can include the user's location, if he/she is inside or outside, his/her GPS position, if he/she is at home, at work or in transportation. Moreover, it can also include information about his/her surroundings. For example, it would be interesting to know if he/she is alone or in a crowded environment, if he/she is in a noisy environment or what the temperature and the illumination conditions of his/her environment are.
- **The user monitoring information.** The interaction between an interactive system and a user can be monitored and then reported into his/her user model. For instance, it can be used to gather the user's interaction habits and interests.

2.2 Machine learning for personalized user interfaces

In the field of adaptive user interfaces, machine learning is mostly used for content selection. These solutions refer to recommender systems (RSs). RSs are software tools and techniques providing suggestions for items that are most likely of interest to a particular user such as an item to buy or a piece of music to listen [24]. For instance Muramatsu and Billsus [20] propose a Bayesian classifier to learn which kind of web pages a user will like or dislike for a given topic. As well, Schwarzkopf also introduces a bayesian network in order to create a personalized web portal page with links and items that match the user's interests [25]. In that category, Maes and Kozierok [16] also propose a reinforcement learning solution in order to train an interface agent that can assist a user in a scheduling application and in an electronic mail application. The trained agent can make interaction suggestion and can also automate an action. A complete survey of recommender systems has been introduced by Ricci et al. [24]. Today, recommender systems based on machine learning are highly developed and can be found on a large variety of websites such as e-commerce websites (Amazon), social networks (Facebook, Twitter) and search engines (Google, Bing).

Other uses of machine learning for personalized user interfaces also include user tutoring such as guidance, task support such as auto-completion, layout personalization, and personalized item differentiation [22].

Most of these work focuses on 2D user interfaces (websites and standalone applications) and as such applications are mainly based on the WIMP (Window, Icon, Mouse, Pointer) interaction paradigm or on tactile 2D interactions, only a few works are interested in adapting the interaction techniques to the target users. As such a standard

does not really exist for 3D user interfaces, the variety of interaction techniques that can be used is much wider. This diversity creates the challenge of finding the most adapted interaction techniques according to the target context of use (use case, hardware, users, etc.). For now, only a few works focused on using machine learning to deal with these issues.

2.3 Interaction techniques adaptation in 3DUIs

Solutions for adapting interaction techniques to the users are not common in the field of 3DUIs. In that category, Octavia et al. [18] propose a conceptual framework for the creation of 3DUIs with interaction techniques adapted to user preferences. First, this framework is composed of a user model that includes the user's interaction patterns, interests, and characteristics. They define three levels of information included in this model. The general user model includes basic information used for adaptation to all users. The group user model provides specialized information to be applied to a group of users. For instance, group user models can be used in order to detect the user preferences according to his/her profile. The individual user model gives detailed information about one particular user. Second, the framework is composed of an adaptation engine that modifies the interaction techniques according to the user model. Interaction techniques can be replaced, disabled or enhanced with other modalities. They have investigated the creation of these different levels for a selection task in a VE. Two selection techniques were compared: a ray-based interaction technique and a bubble cursor. Physiological data were measured in order to evaluate user frustration, user experience, and mental workload. They did not find any trend for the creation of group user models. Nevertheless, they have shown in [19] that employing general and individual user models can increase users' performances and can decrease users' frustrations. The general user model was created according to the scene parameters and an individual user model was created for each user after a first experiment with the two interaction techniques. Such approaches do not really give the possibility to adapt an application to the preferences of a new user. That is why more work should be done to create a solution that can predict the preferences of a new user according to the collected information about him/her. An interesting perspective of work has been proposed by Wingrave et al. [28]. They address interaction techniques exchangeability and intrinsic parameters tuning from users with two different selection techniques: Ray-casting and Occlusion selection. During the experiment, they did not find any trend, each user had his own preferences. That is why they propose a perspective of work that aims at correlating the user's profile and behaviour with his/her preferences. This approach consists in using "nuances" in an inverse reinforcement learning process to predict user behaviors. A nuance is defined by Wingrave et al. [27] as *"A repeatable action the user makes in their interactions in an environment, intentional or not, that are highly correlated with an intended action but not implicit in the interaction metaphor"*. These different work did not succeed to build a function that could associate a new user to his/her adapted interaction. That is why we propose the machine learning approach to try to learn this function. This approach requires the creation of a dataset that contains the association between users and their performances and preferences for three different 3D selection techniques.

3 Dataset creation

Our objective is to be able to train a machine learning algorithm that can predict the most adapted interaction technique to a particular situation according to the user information. That is why we propose to build a dataset in order to validate such an approach. To build such a dataset, we have developed a 3DUI with a classic 3D selection task on two different VR hardware configurations. Users experimented with this application on these two configurations and with three different interaction techniques. Using two different hardware configurations can allow us to check if it is possible to use the prediction model for different kinds of hardware contexts. Subjective preferences and performance results were recorded during each session. Indeed, in order to determine what is the most adapted technique to a particular situation we consider two different strategies, first we can select the one which obtains the best subjective results (the "preferred one"), second, we can select the one with which a user performs the best (the "best performing" one). Regarding user modeling, we explored two different approaches. First, we asked users different questions about their profile such as their age, gender, experience with VR, etc. Second, as asking such questions could raise confidentiality issues for end-users, we also propose to model each user with a simple selection pre-test. A possibility could be to let the user try the three interaction techniques and let him choose the one he prefers or choose for him the one with which he performed the best. Depending on the number of techniques, this approach would require time and the result could depend on the order of the interaction techniques. Therefore, we propose a pre-test based on a simple 2D selection task. We hope that a machine learning algorithm will be able to associate the results of this pre-test with the most adapted interaction technique for our 3D selection task.

Hardware Each user performed the same selection on two different hardware setups that differ by the level of immersion that they provide. The main difference between the two platforms was the display used as the control of the interaction techniques was approximately the same on the two platforms with a 6-DOF controller. First, the zSpace that is shown in Figure 1a is composed of a 6-DOF tracker with buttons (the stylus) and a 24-inches co-localized stereoscopic display. Second, we used a setup based on an Oculus Rift DK2 HMD and a Razer Hydra as shown in Figure 1b. The Razer Hydra is composed of two 6-DOF controllers with buttons. Only one of them was used.

Participants Our goal was to collect as much data as we could because the efficiency of a machine learning algorithm mainly depends on the size of the training dataset. That is why our experimental group consisted of 51 subjects aged from 23 to 57 (age: $M=34.8$, $SD=10.24$). There were 38 males and 13 females. The subjects had various backgrounds, some of them were not working in the technology industry. Most of them had little knowledge about VR.

Interaction techniques We compared three different 3D selection techniques, taken from the state of the art, that allow close and distant objects selections. For each technique, two-color feedbacks were provided to the user. A pointed object was colored in yellow, and a selected one in red. These techniques were the following ones:

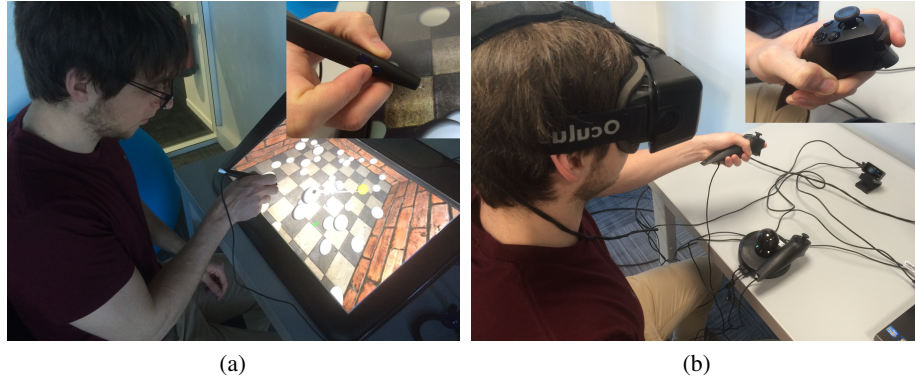


Fig. 1: The two device setups used for our user preferences study (a) The first setup is zSpace composed of a stereoscopic screen with head tracking and of a 6-DOF tracked stylus. (b) The second setup is composed of an Oculus Rift DK2 and a Razer Hydra.

- **3D Ray-Casting**: this first interaction technique is based on the ray-casting metaphor as described by Bowman et al. [2]. A straight ray that comes from the user’s hand and intersects a scene object for selection as described in Figure 1a. The ray base is controlled in position and rotation with the values given by the stylus tracker or by one Razer Hydra controller. A button was used for selections.
- **3D Bendcast** this technique is very similar to the first one. It is also controlled with the stylus or by one Razer Hydra controller. The technique is described in Figure 3(b). Our implementation is similar to the one described by Cashion et al. [4], the Bendcast technique automatically bends to the closest target in order to help the user during the selection task. One difference in our implementation is that the ray bends only if the distance between the ray and the object is under a threshold. This threshold is chosen according to the size of each target.
- **Occlusion**: as described by Pierce et al. [21], the occlusion selection technique consists in placing the tip of your finger between your eye and the object you want to select. The occluded object is the one selected. In our implementation, the stylus extremity or one Razer Hydra controller is used to drive the occluder and a button is also used to confirmed selection. The eye used by the technique corresponds to the user’s master eye. The technique is described in figure 3(c).

Experimental protocol The first step of the experimental protocol consisted in doing a 2D selection task based on Fitts’s law. A view of a participant performing this test is given on the right of Figure 2. The Fitts law describes the relationship between the target distance, width and time needed for a selection task. Fitts’s law can be used in order to compare the performances of pointing devices by analyzing the throughput. Our

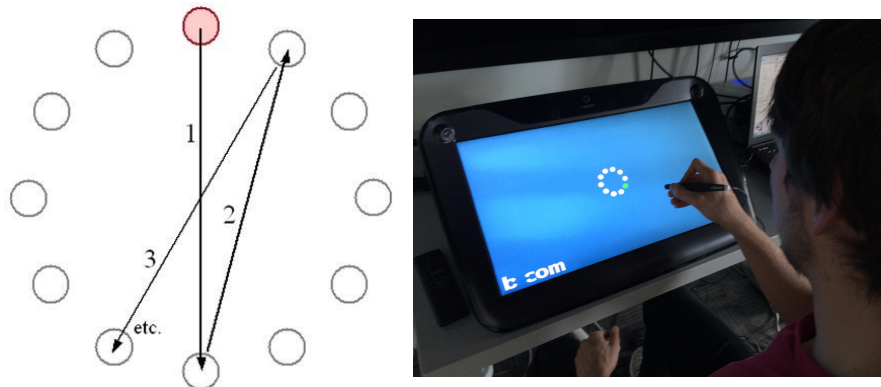


Fig. 2: ISO 9241-9 multi-directional pointing task. The user first selects the highlighted circle, then the next circles to select follow the pattern drawn with the arrows..

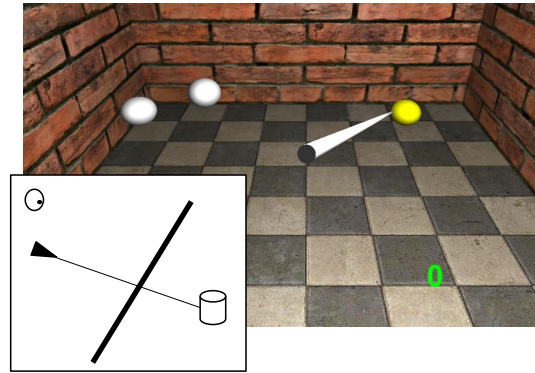
Fitts's law based tests follow the ISO9241-9 standard and the recommendations given by Soukoreff and Mackenzie [26].

This first part of the experiment was performed by each user on the zSpace. No stereoscopy or head tracking were used. The targets were simply displayed in 2D on the screen of the device. The 6-DOF stylus and one of its buttons were used to control a 2D cursor to perform selections. The cursor position corresponded to the intersection between the stylus and the screen. Therefore, to move it the user had to target the screen as shown in Figure 2. Two-color feedbacks were provided. The disk to select was colored in green and when selected (clicked) the disk was colored in red. During this part of the experiment this multi-directional pointing task was performed three times by each participant with three different IDs (difficulty increases over time):

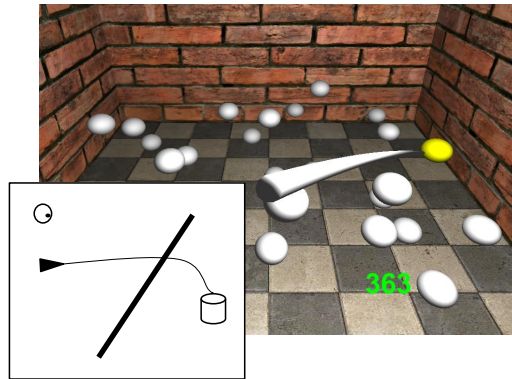
- disk diameter = 40 pixels ; distance = 200 pixels ; $ID = 2.58$
- disk diameter = 20 pixels ; distance = 200 pixels ; $ID = 3.46$
- disk diameter = 20 pixels ; distance = 500 pixels ; $ID = 4.7$

The goal of this first test is to get a first result on the interaction skill of each user for a basic task. As this 2D selection task is done with a 6-Dof controller (the zSpace stylus) it can also be a clue to the 3D interaction capabilities of the user. That is why we propose to try to use the monitoring information of this task in order to construct a machine learning based prediction model for choosing an adapted 3D selection technique for a 3D selection task.

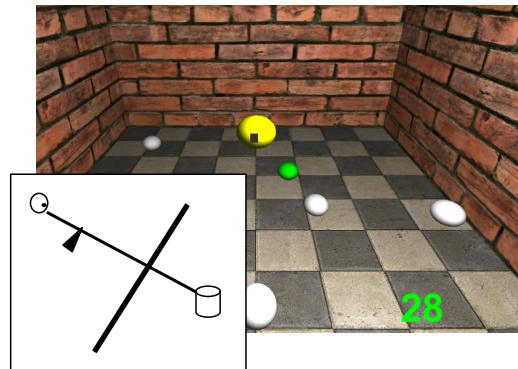
The second part of the experiment proposed a series of 3D selections to perform with the two interaction setups and with the three interaction techniques presented before. The 3D selection task was prototyped to be as generic as possible in order to allow us to use the results for other kinds of virtual environments and objects. In this part of the experiment, participants were placed in a virtual environment at scale one just in front of



(a)



(b)



(c)

Fig. 3: The three interaction techniques compared in our user preferences study. (a) A 3D Ray-Casting interaction technique. The position and the orientation of the ray directly follow the 6-DOF controller. (b) A 3D Bendcast interaction technique. This technique is similar to the first one. Here, the ray bends in order to automatically sticks to a target if it is close enough (under a threshold). (c) The Occlusion selection technique. Here, the user has to hide an object with his controller extremity in order to point it. As shown, a small black cube is displayed in the VE at this extremity.

a box containing different spheres to select as shown in Figure 3. These spheres were all the same fixed size: 3cm of diameter.

For this part of the experiment, some users started with the Oculus Rift while other users started with the zSpace. In order to pass to the next setup, the user had to perform all selections in four scene configurations and with the three interaction techniques. All the users did not try the techniques in the same order. Nevertheless, the same order was applied to both setups for one particular participant. We applied a counterbalanced design: $2 \text{ device setups} * 6 \text{ technique orders}$. For each technique, the process was the same. The users had to select the spheres in four scene configurations. For each configuration, the user had to perform 10 selections. The sphere to display was colored in green as illustrated in Figure 3(c). These scene configurations were depending on two parameters: two conditions of density (low and high) and two conditions of movement (mobile, stationary). For each interaction technique, the scene configuration order was always the same: low density/stationary spheres, high density/stationary spheres, low density/mobile spheres, and high density/mobile spheres. The goal was to increase the difficulty progressively. The instruction given to each user was to select the spheres as fast as possible and to avoid mistakes. In order to challenge them, we displayed a score about their performances. This score took into account the selection times and was decreased when a wrong selection was made.

Collected Data First, in order to build the user profile of each participant that is integrated into the user model, each of them was asked to fulfill a preliminary questionnaire. The user profile consists of a set of parameters directly given by the user in order to define him/her as a person. The following pieces of information were requested:

- gender,
- handedness,
- age,
- master Eye,
- visual problems (Yes/No),
- experience with virtual reality (Yes/No) (Hours per Week),
- experience with video games (Yes/No) (Hours per Week),
- experience with computers/tablets (Hours per week).

During the 2D selection pre-test, all the actions of the user were collected. Indeed, for instance, we collected the selection times, the cursor movements relative to the straight path between two targets, the number of clicks outside a disk or on a wrong one, the distance from a click to the center of click, etc.

For the 3D selection tasks, in order to get the preferences of all users, they had to fulfill different subjective questionnaires at different steps of the experiment. In these subjective questionnaires, participants had to grade the three techniques using a Likert-scale, from 1 (very low appreciation) to 7 (very high appreciation) according to one criterion: global appreciation. The goal was just to record which technique they were preferring. A first grade had to be given to an interaction technique after having performed all selections for one scene configuration. Then, after having finished all selections on the first device setup, users were asked to give new grades to each technique, a global one and one for every scene configuration. At the end of the experimentation,

the user had to give again different global grades for each technique on each device, and also to each technique for each scene configuration for each device setup. During this step, the participants did not have access to the grades they gave earlier in the experiment. Each time, the given grades are final and cannot be modified later in the experiment. The goal of this last step was to allow the user to make a comparison between the two hardware configurations. The user's actions and performances were also collected during this part of the experiment: selection times, the number of errors, the quantity of movement, speed, etc.

4 Prior Analysis

A prior analysis of the results obtained during the dataset creation is required in order to determine some guidelines for the training of our prediction algorithms. First, we propose to compare the preferences and performances of the three interaction techniques in order to determine if an adaptation mechanism is really required in our situation. Second, we want to verify the consistency between the results on the two devices setups. Indeed, it is necessary to know if the target device must be taken into account in the prediction model. Last, we need to verify the consistency between the preferred techniques and the best-performing ones in order to know if our models must be trained to a particular goal (preference or performance) or if they can be completely generic.

Global performances and preferences First, we have analyzed the preferences scores and the performance data obtained for each technique for the two device setups. The goal of such an analysis is to identify the technique that has to be selected by default if we do not know anything about the end-user. This can be compared to the general user model presented in [19]. This general user model consists in choosing the same interaction technique for each user. This interaction technique corresponds to the one that has been chosen by the majority of users. This general model is compared to our prediction algorithms in Section 5. For now, our analysis of the results does not separate the performances and the preferences according to the different scene configurations. Indeed, we have added the preferences scores and the performance scores of each technique over the selections performed for the four scene configurations proposed. For each technique and for each setup two different global grades between one and seven were given by the participants. The first one was requested after having finished the selections on one device setup. The second one was requested at the end of the experiment. These two grades have been added in order to have the global appreciation of the user for each technique. For the performances, we have used the scores that we computed at runtime that correspond to a combination between the selections times and the number of errors. The more this score is important, the better the performances were. As this paper mainly focuses on interaction technique personalization we do not detail all the statistical tests that we performed. In Table 1, we present how many times each technique has been ranked at the first, second and third places for the preferences and performances on the two hardware configurations. For some users, different techniques obtained equal preferences scores and therefore have been ranked at the same place. This table suggests that users globally preferred interacting with the 3D Ray-based interaction

technique and with the 3D Bendcast technique. Globally only significant differences were found between the occlusion technique and the two other ones. As in most cases presented here the 3D Ray-based interaction technique and the 3D Bendcast got similar results we can say that in general case it is complicated to choose a technique that will satisfy all users. Moreover, even if the occlusion selection technique globally got the worst results, it is still represented by 10 to 20% users depending on the chosen criterion. These assessments justify the need for user adaptation more advanced than this kind of general user model.

Consistency between hardware configurations Then, we verified the consistency between the performances and the preferences on the two device setups. To make a global comparison of the preferences and the performances on the two configurations, we created a global preference score on each device configuration and a global performance score. Both are the addition of the preferences and the performance scores (respectively) of all techniques for a given user. Globally, users better performed on the zSpace than with the Oculus Rift. One-way analysis of variance (ANOVA) showed a significant main effect of the device setup on the performances ($F_{1,100} = 131.0, p < 0.005$). In the same way, the three interaction techniques have globally obtained better grades on the zSpace than with the Oculus Rift. One-way analysis of variance (ANOVA) showed a significant main effect of the device setup on the grades ($F_{1,100} = 33.57, p < 0.005$). From these two results, we can say that participants significantly preferred to interact on the zSpace and significantly performed better on this same device setup for our selection task. Therefore, we can say that even with the similarities between the two device setups (the same kind of 6-DOF controller was used to control the interaction techniques), the performances and global appreciations were different. To continue, we also wanted to know if the preferred interaction techniques or the ones that got the best performances are different on the two device setups. Therefore, we have computed the number of times the preferred interaction technique or the one that performed the best has changed from the first device setup to the second one. In the case of an equality of grades in the first place, multiple techniques were considered as the preferred one. We obtained the following results:

- Regarding the preferences grades, 10 participants (19,6%) did not prefer the same interaction technique on both device setups. For these 10 participants, the grade of a preferred technique on a given device setup had decreased by an average of 14.70% on the other device setup.
- Regarding the performance scores, for 33 participants (64.7%) the interaction technique that obtained the best performance was not the same on the two device setups. For these 33 participants, the score of a technique that performed the best on a given device setup had decreased by an average of 8.94% on the other device setup.

From these results, we can say that the preferred interaction technique or the one with which a user performs the best will not always be the same on all devices setup. Therefore, we can say that taking into account the target hardware configuration is a requirement that our prediction algorithms must take into account.

Oculus Rift							
Preferences				Performances			
	1st	2nd	3rd		1st	2nd	3rd
Ray	27	21	3		22	20	9
BendCast	26	21	4		22	21	8
Occlusion	9	10	32		7	10	34

Zspace							
Preferences				Performances			
	1st	2nd	3rd		1st	2nd	3rd
Ray	19	30	2		17	24	10
BendCast	35	11	5		24	17	10
Occlusion	11	12	27		9	10	31

Table 1: Number of times that each technique has been ranked as first, second or third according to the preference or the performance criteria on the two hardware configurations

Consistency between preferences and performances Last, for each user, we have verified on each hardware setup if the preferred interaction techniques matched the best performing one. To do so, we have computed on each setup the number of times the technique that obtained the best performance did not correspond to the one that the user preferred. As well, in the case of an equality of grades in the first place, multiple techniques were considered as the preferred one. We obtained the following results:

- With the Oculus, for 25 participants (49%), the preferred interaction technique did not correspond to the one that performed the best. For these 25 participants, the average difference between the performance score of the technique that performed the best and the performance score of the preferred technique was 8.22%. The average difference between the grade of the preferred technique and the grade of the technique that performed the best was 22.32%.
- On the zSpace, for 21 participants (41%), the preferred interaction technique did not correspond to the one that performed the best. For these 21 participants, the average difference between the performance score of the technique that performed the best and the performance score of the preferred technique was 6.59%. The average difference between the grade of the preferred technique and the grade of the technique that performed the best was 21.64%.

For the two device setups, we obtained similar results. Approximately one on two users did not prefer the same technique that the one with which he/she performed the best. The average differences that we computed are important enough to say that a choice must be made when selecting a technique for a user. In a lot of cases, there will be a technique that will maximize the user's performances and another one that will maximize the user's preferences. A possibility could also be to select a technique by making a compromise between performances and preferences. Our prediction algorithms must take into account the possibility to make this choice.

5 Interaction Technique choice with Machine Learning

At this point, we have obtained a dataset that includes the correspondences between each user profile and each user results from the 2D pre-test with his/her preferred technique and his/her best performing one on each hardware configuration. As detailed, our goal is to determine if from this dataset we could automatically predict the interaction technique that would correspond to a new user.

From this data, we can train a supervised machine learning algorithm. In order to create a mapping function, supervised learning algorithms are trained on labeled training data that include examples composed of an input object (a vector) and a label (the desired output). To do so, a loss error is minimized between the labels predicted by the function and the initial real labels. From this training, a function that generalizes the problematics is created and in the ideal case, this function will be able to predict the label of an unseen example. To evaluate the prediction robustness and accuracy of such method datasets are commonly separated into two parts. First, the training dataset on which the algorithm is trained. Second, the testing dataset is composed of unseen points, not present in the training set on which the generalization ability is evaluated. In our scenario, we have used the ratio *Good predictions/All predictions* to quantify the accuracy of our created models.

Here, we use a Support-Vector-Machine (SVM) classifier as a supervised learning algorithm. An SVM classifier aims to construct from training examples a set of hyper-planes in a high dimensional space that separates the data into two classes (positive or negative) [1]. Recent approaches also exploit SVMs to deal with multi-classes problems. This is the kind of approach we use here as our data are labeled with three possible classes: the three interaction techniques (3D-Ray, 3D-Bendcast or Occlusion). Our implementation is based on LIBSVM [5] that combines multiple binary classifiers in order to create a multi-classes classifier. Our choice is strongly motivated by the ability of SVMs to deal with a small amount of training data unlike other algorithms such as neuronal networks and random forests. Indeed, our database only contains 51 examples, which would not be enough for these classes of algorithms.

Each of our examples can be formulated as a pair $\{x_i; y_i\}$ in $\mathbb{R}^n \times \{0, 1, 2\}$. Here, x_i corresponds to the input feature vector and contains information extracted from the profile or from the 2D preliminary tests. In all cases, each value contained in x_i is normalized. Binary data such as "handedness" or "experience with VR" are set to 0 or 1. The dimensions of the different feature vectors used are kept fixed during the learning process as we use a linear kernel during the SVM optimization. y_i corresponds to the label, which is the interaction techniques associated with the profile (the preferred one or the one with the best performance), 0 corresponds to the 3D-Ray technique, 1 to the 3D Bendcast technique and 2 to the Occlusion technique. For a given user, in case of equality between the qualitative grades obtained by multiple techniques, only one technique has been selected as the preferred one according to performance ranking. After multiple iterations of tests, the following feature vectors have been selected in order to describe each user in our different tests. For the profile information we use the following

feature vector:

$$x_i = \{ \text{Wear glasses (binary)}, \text{VR experience (binary)}, \\ \text{Hour/week VR (real)}, \text{Video game experience (binary)}, \\ \text{Hour/week Videogame (real)} \} \quad (1)$$

For the information extracted from the 2D pre-test we use the following feature vector:

$$x_i = \{ \text{Number Click Outside (real)}, \text{Total Selection Time (real)}, \\ \text{Min Selection Time (real)}, \text{Click Mean Distance Center (real)}, \\ \text{Click Min Distance Center (real)}, \text{Click Max Distance Center (real)}, \\ \text{Distance Max Straight Path (real)} \} \quad (2)$$

As detailed in Section 4, the best performing technique does not always match the preferred one. As well, we saw that an adapted technique can change from one device to another one. To deal with this issue we have tried two solutions. First, we have tried to include the hardware configuration and the criterion of adaptiveness (performance or preference) as binary parameters. Second, we have tried to build different classifiers for each situation. Regarding the preliminary accuracy results, we chose the second option. Therefore, we propose to build multiple classifiers to predict the preferred technique on the Oculus Rift, the best performing technique on the Oculus Rift, the preferred technique on the zSpace and the best performing technique on the zSpace. For all cases, we compare the results obtained for the SVM with the result with a simple "majority" method. This method can be compared to the general user model proposed by Octavia et al. [19]. It consists in selecting for a new user the technique that has performed or has been preferred the most by the participants on one setup. Our goal is to obtain better ratios with our machine learning approach.

In order to determine the prediction accuracy of our models, in all cases, we separate our data into two datasets: 40 examples in the training set and 11 examples in the testing set. During the training step, we have performed a k-fold cross-validation in order to tune one parameter of the SVM. This parameter named C tells to the SVM optimization problem how much it has to avoid misclassifying each training example. During this cross-validation, the metric to optimize in order to choose the best value of C is the ratio *Good predictions/Number of predictions*. Because of the small size of our training data, the generalization ability is a very challenging task. To enhance and guaranty this ability, we have decided to repeat the process 1000 times where each time the training set and the testing set are fulfilled randomly. The obtained results correspond to the average value of the ratio of good predictions on the testing set and the standard deviation for this same value obtained for the SVM classifiers evaluated on the 1000 separations. The results obtained for the Oculus Rift configuration are detailed in Table 2 and the ones for the zSpace in Table 3. These results suggest that in most cases our classifiers have a better prediction rate than the general user model. However, the gains of using our method are relatively low. Moreover, with the pre-test features on the zSpace, we can see that the results are worse than the general user model.

	Preferred Technique	Best Performing Technique
General User Model	49.02%	43.14%
SVM with profile features	M=57.17% SD= 12.99%	M=52.95% SD= 13.85%
SVM with pre-test features	M=55.60% SD=13.97%	M=52.95% SD=13.03%

Table 2: Comparison of the prediction ratios (good predictions / false predictions) obtained with our different classifiers for the data obtained on the Oculus configuration.

	Preferred Technique	Best Performing Technique
General User Model	47.05%	47.05%
SVM with profile features	M=52.36% SD= 13.94%	M=35.98% SD= 11.79%
SVM with pre-test features	M=39.12% SD=11.79%	M=38.47% SD=12.97%

Table 3: Comparison of the prediction ratios (good predictions / false predictions) obtained with our different classifiers for the data obtained on the zSpace configuration.

6 Discussion

Thanks to the creation of this dataset we can confirm that even with a simple task and with three "state of the art" interaction techniques, the distribution of preferences and performances among users for these techniques was really dispersed. Such results can be found in most studies that compare different 3D interaction techniques. Even, if significant differences are found between interaction techniques, there will always be some users that have different results from the others. Such observations highlight the need for personalization systems in the field of 3D user interfaces.

To continue, this experiment shows that there are two important requirements when building such personalization systems for 3D user interfaces. First, as such interfaces can be run on multiple kinds of hardware setups, we have demonstrated that this setup has an impact on the preferred and best performing interaction techniques for a target user. This hardware information must be taken into account during the choice of the interaction technique. Second, we saw that the preferred interaction techniques and the ones with which a user performs the best do not always match. Thus, during the choice of the most adapted interaction technique it must be specified which one has to be selected: the

preferred one or the best performing one. Such a parameter could be controlled by the application developer.

As detailed, in most cases, we can see that our machine learning approach can provide better ratios compared to a classic "majority method" or to a random choice. Better results were obtained on the Oculus Hardware than on the zSpace setup. However, the improvements of ratios of good predictions compared to the "Majority" method are not important enough to really apply our SVMs in their current states in order to detect the best-suited interaction technique for a new user. Indeed, even with our better ratio, 43% of users would not get the most adapted technique. Moreover, we can notice that better results were obtained on the Oculus Setup with the pre-test features. This is quite surprising as this pre-test was performed on the zSpace. This result suggests that such a pre-test can be done in a different hardware setup than the one used for the main task. Additional work must be performed to confirm this statement. Globally, these results suggest that our approach can be a good way to detect a well-suited interaction technique for a target, but it requires additional work to improve its accuracy results. We think that the small size of our dataset was the main reason for these results and that with the construction of a larger dataset we would be able to get better results with our approach.

Last, for now, our approach suffers from generalization. Indeed, it can only detect the most adapted technique between three interaction techniques for a selection task and for the two given setups. Other experiments would be needed if we want to create other SVMs to predict other kinds of interaction techniques such as navigation and manipulation techniques. As well, the choice of an adapted interaction technique in other selection scenarios must be validated if we really want to demonstrate the efficiency of our approach. We think that it could be possible to use SVMs to predict global characteristics from the user profile or from a preliminary test and then match these characteristics with the properties of some interaction techniques. For example, we could detect automatically a novice user with low skill and automatically propose him/her 3D menus with bigger icons and an assisted selection technique such as the 3D Bentray.

7 Conclusion and Future Work

To conclude, in this paper we propose a machine learning approach in order to determine which 3D selection technique will be the most suited to a target user according to his/her profile or to his/her results to a 2D selection pre-test. The requirements of this approach and its validation have been performed with the creation of a dataset by collecting data from 51 users in a 3D selection task, with three different interaction techniques and two hardware setups.

Accuracy results suggest that our approach has globally a better chance to find the most suited technique to a target user compared to the "majority method" where the same interaction technique is selected for everyone. However, for now, these results are not good enough to be sure that each user will get the most adapted technique. As machine learning algorithms such as SVM require a large amount of data during the training phase, we guess that increasing the size of the dataset could be a solution to improve our results. This is required if we want our solution to be possibly integrated in a framework for the creation of adaptive 3D user interfaces.

Therefore, our future work will focus on increasing the size of our dataset in order to try to improve the accuracy of our results. Then we also want to validate the approach on other kinds of 3D interaction techniques. Last, we plan to investigate a third approach to construct the user model: this approach would consist in using the monitoring information of the users while they are interacting with the system.

8 Reference to the official publication

Springer Nature Switzerland AG 2019

P. Bourdot et al. (Eds.): EuroVR 2019, LNCS 11883, pp. 1–19, 2019.

The final authenticated version is available online at:

https://doi.org/10.1007/978-3-030-31908-3_3.

References

1. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory. pp. 144–152. ACM (1992)
2. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA (2004)
3. Calvary, G.: Plasticité des Interfaces Homme-Machine. Ph.D. thesis (2007), thèse Habilitation à Diriger des Recherches préparée au Laboratoire d’Informatique de Grenoble (LIG), Université Joseph Fourier.
4. Cashion, J., Wingrave, C., LaViola, J.: Optimal 3d selection technique assignment using real-time contextual analysis. In: 3D User Interfaces (3DUI), 2013 IEEE Symposium on. pp. 107–110 (March 2013)
5. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) **2**(3), 27 (2011)
6. Chittaro, L., Ranon, R., Artificial, I.S., Realities, V.: Dynamic generation of personalized VRML content: a general approach and its application to 3d e-commerce. In: In Proceedings of Web3D 2002: 7th International Conference on 3D Web. p. 145–154. Press (2002)
7. Dachsel, R., Hinz, M., Pietschmann, S.: Using the AMACONT architecture for flexible adaptation of 3d web applications. In: Proceedings of the Eleventh International Conference on 3D Web Technology. p. 75–84. Web3D ’06, ACM, New York, NY, USA (2006). <https://doi.org/10.1145/1122591.1122602>, <http://doi.acm.org/10.1145/1122591.1122602>
8. Dragicevic, P., Fekete, J.D.: Input device selection and interaction configuration with ICON. In: People and Computers XV—Interaction without Frontiers, p. 543–558. Springer (2001)
9. Fischer, G.: User modeling in human–computer interaction. User modeling and user-adapted interaction **11**(1-2), 65–86 (2001)
10. Hand, C.: A survey of 3d interaction techniques. In: Computer graphics forum. vol. 16, pp. 269–281. Wiley Online Library (1997)
11. Kobsa, A.: User modeling: Recent work, prospects and hazards. Human Factors in Information Technology **10**, 111–111 (1993)
12. Kobsa, A.: Supporting user interfaces for all through user modeling. Advances in Human Factors/Ergonomics **20**, 155–157 (1995)
13. Kobsa, A.: Generic user modeling systems. User modeling and user-adapted interaction **11**(1-2), 49–63 (2001)

14. Lacoche, J., Duval, T., Arnaldi, B., Maisel, É., Royan, J.: Providing plasticity and redistribution for 3d user interfaces using the d3part model. *Journal on Multimodal User Interfaces* **11**(2), 197–210 (2017)
15. Lindt, I.: Adaptive 3D-User-Interfaces. Ph.D. thesis (2009)
16. Maes, P., Kozierok, R.: Learning interface agents. In: *Proceedings of the Eleventh National Conference on Artificial Intelligence*. pp. 459–464. AAAI'93, AAAI Press (1993)
17. Mohri, M., Rostamizadeh, A., Talwalkar, A.: *Foundations of machine learning*. MIT press (2012)
18. Octavia, J.R., Raymaekers, C., Coninx, K.: A conceptual framework for adaptation and personalization in virtual environments. In: *Database and Expert Systems Application, 2009. DEXA'09. 20th International Workshop on*. p. 284–288 (2009)
19. Octavia, J., Raymaekers, C., Coninx, K.: Adaptation in virtual environments: conceptual framework and user models. *Multimedia Tools and Applications* **54**(1), 121–142 (2011). <https://doi.org/10.1007/s11042-010-0525-z>, <http://dx.doi.org/10.1007/s11042-010-0525-z>
20. Pazzani, M.J., Muramatsu, J., Billsus, D., et al.: Syskill & webert: Identifying interesting web sites. In: *AAAI/IAAI*, Vol. 1. pp. 54–61 (1996)
21. Pierce, J.S., Forsberg, A.S., Conway, M.J., Hong, S., Zeleznik, R.C., Mine, M.R.: Image plane interaction techniques in 3d immersive environments. In: *Proceedings of the 1997 symposium on Interactive 3D graphics*. pp. 39–ff. ACM (1997)
22. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D.: Web usage mining as a tool for personalization: A survey. *User modeling and user-adapted interaction* **13**(4), 311–372 (2003)
23. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: *Recommender systems handbook*, pp. 1–35. Springer (2011)
24. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B.: *Recommender systems handbook*. Springer (2015)
25. Schwarzkopf, E.: An adaptive web site for the um2001 conference. In: *Proceedings of the UM2001 workshop on machine learning for user modeling*. pp. 77–86 (2001)
26. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International journal of human-computer studies* **61**(6), 751–789 (2004)
27. Wingrave, C.A., Bowman, D.A., Ramakrishnan, N.: A first step towards nuance-oriented interfaces for virtual environments (2001)
28. Wingrave, C.A., Bowman, D.A., Ramakrishnan, N.: Towards preferences in virtual environment interfaces. In: *Proceedings of the Workshop on Virtual Environments 2002*. p. 63–72. EGVE '02, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2002), <http://dl.acm.org/citation.cfm?id=509709.509720>